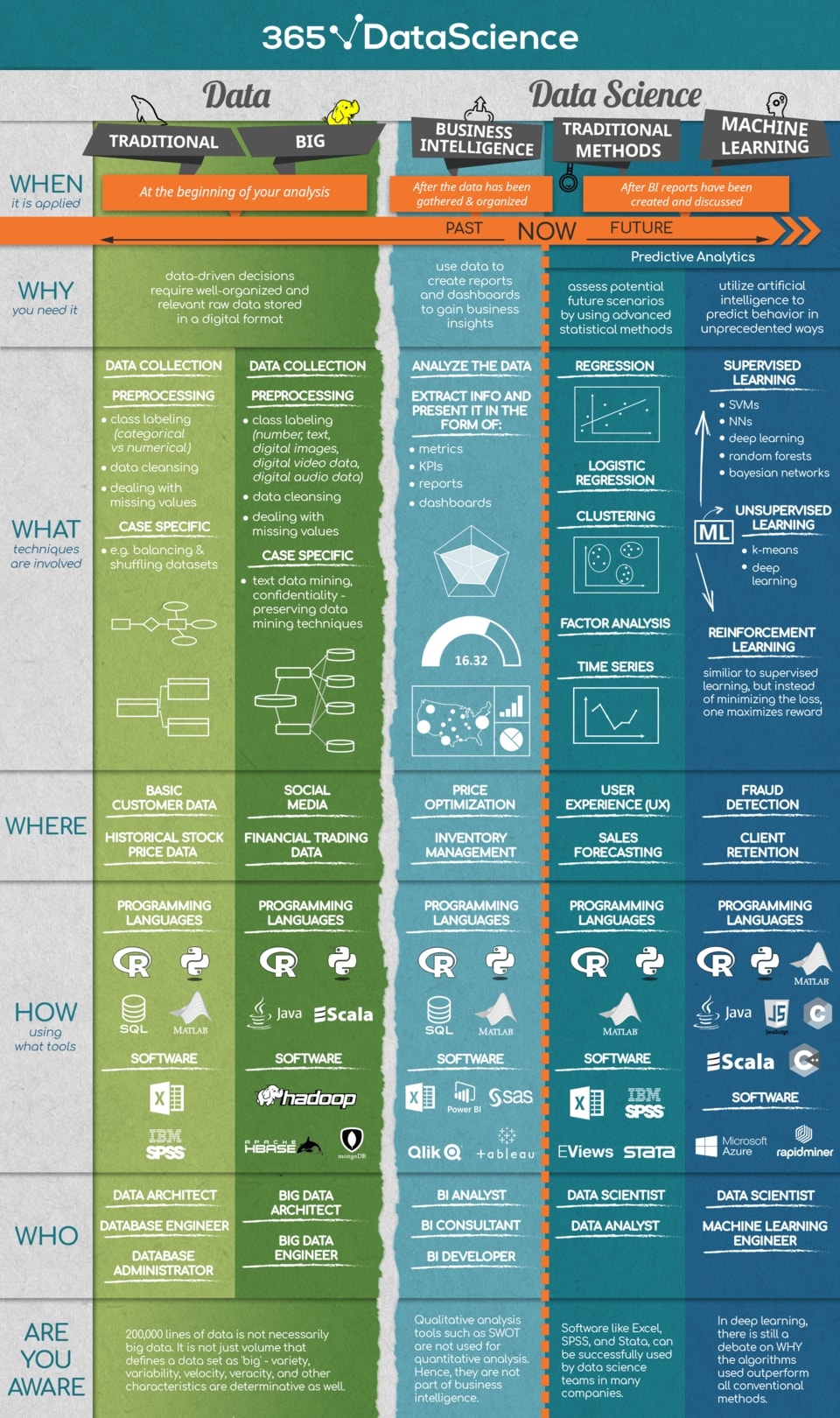
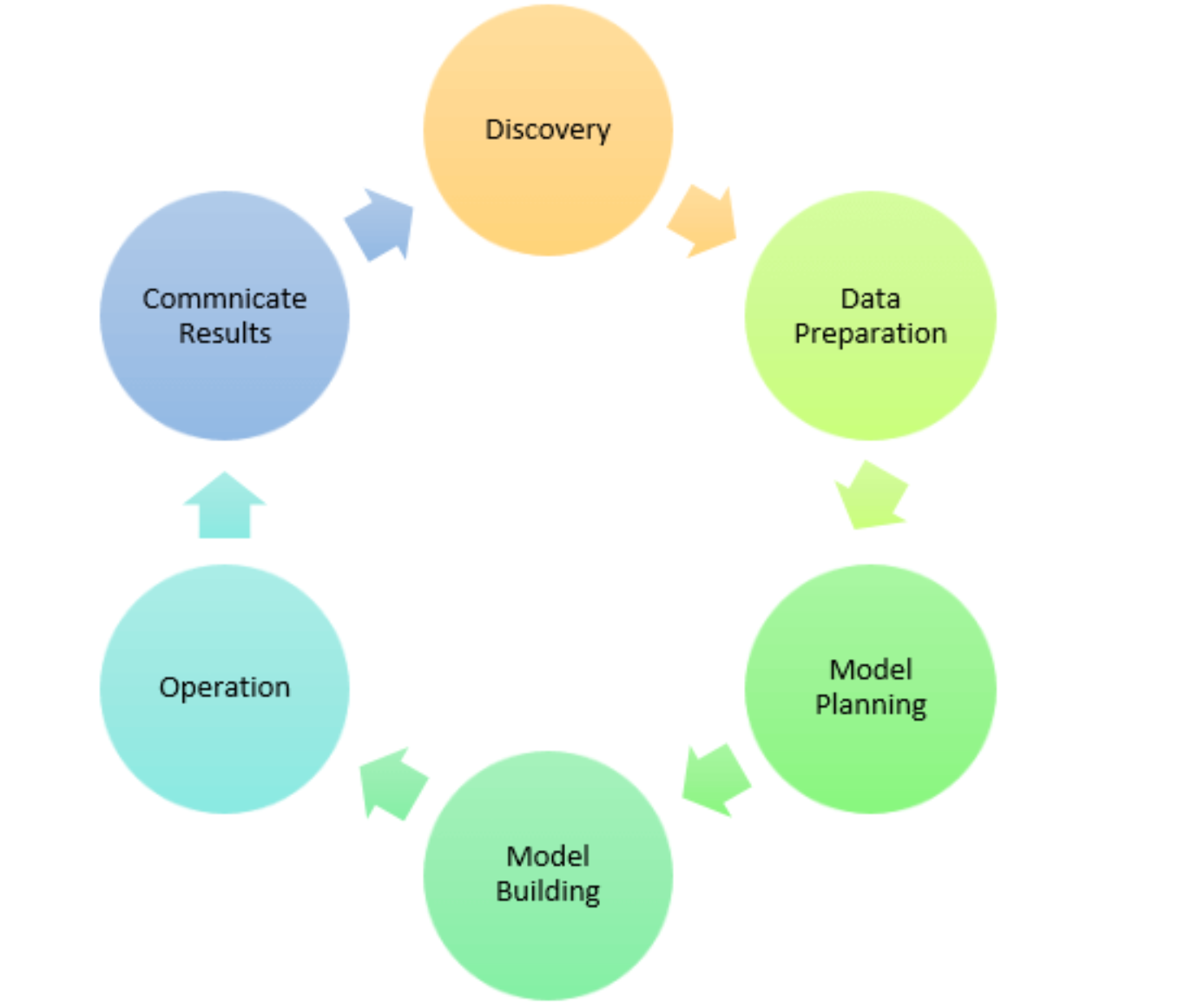
**What is data science?** Data science is an inter- disciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insight from many structural and unstructured data. **What is data analysis?** Data analysis is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision making. **What is decision analysis?** Decision analysis is the discipline comprising the philosophy, methodology and professional practice necessary to address important decisions in a formal manner. Let us understand when, why, what, where, how, and who of data science in the below representation.



There is a lot of information in the above diagram but to keep it crisp to the point. As we have the gathered data from different sources. By using the gathered data, we can analysis and decide a future prediction which should be relevant based on the data source. Once the decision is made, we can follow to the analysis part. we need to create reports to understand the business’s performance better. Now is the time to concentrate on the predictive analytics, by this analysis we can have the future scenarios and behaviors of the customers. And this helps to improve the future sales and profits.

There are few fundamental processes of data science:

1. Discovery: This process involves the data that is gathered from all the sources like both internal and external sources. Which helps us to answer all the business questions. This data can be in logs from webservers, data gathered from social media, census datasets, and data streamed from online sources using API’s.
2. Data preparation: This process helps us to clean the inconsistent data like missing values, blank columns, and incorrect data format. We need to prepare the data before modeling to achieve predictions nearer to accurate.
3. Model planning: In this process, we need to draw the relationship between the attributes to identify the methods and techniques that are needed to perform the operation to get desired predictions. Plan the model to perform by using visualization tools.
4. Model building: In this step, we build the model with the reference of the model planning techniques and methods. Dividing the data set into training and testing, to perform tests.
5. Operation: In this stage, we deliver the final documentation like reports, code, and technical documents with the base model.
6. Communicate results: In this step, the main identifications are to communicate all the findings to the stakeholders. This helps us to decide whether the project is successful or a failure.



**References:**

1. <https://en.wikipedia.org/wiki/Data_science>
2. <https://en.wikipedia.org/wiki/Data_analysis>
3. <https://en.wikipedia.org/wiki/Decision_analysis>
4. <https://365datascience.com/wp-content/uploads/2018/05/365-Data-Science-Infographic.jpg>
5. <https://365datascience.com/defining-data-science/>
6. <https://www.guru99.com/data-science-tutorial.html>